

Etude chimique du collage des cotons

2^e partie : Application de l'analyse des données à une population de 58 cotons

J. Fournier *, J. Gutknecht **, E. Jallas *** et J. Bourely ****

CT 55/100

* Technologiste.

** Chef de la Division de Technologie.

*** Biométricien.

**** Chef du Laboratoire de Chimie des Plantes Textiles.
I.R.C.T.-C.I.R.A.D., B.P. 5035, Montpellier Cedex, France.

RÉSUMÉ

Pour expliquer le collage des cotons dont l'effet préjudiciable à la bonne marche de la filature est certain, 58 cotons de différents degrés de collage mesurés par la méthode de la minicarde ont été étudiés par l'analyse des données : composantes principales, segmentation par les corrélations et segmentation par maximisation de l'écart de variance intergroupe.

Les variables explicatives retenues pour l'analyse sont les suivantes : pH, extrait sec total après macération de la fibre dans l'eau, colorimétrie avant et après étuvage, enfin teneur en différents sucres par chromatographie sur couche mince.

Les résultats obtenus soulignent tout d'abord qu'il faut éviter de confondre systématiquement le collage avec la présence de sucres, puisque les variables les plus liées au collage sont le pH et

l'extrait sec total. La première de ces variables n'est pas du tout liée à la présence de sucres et la deuxième ne l'est que faiblement.

L'analyse en composantes principales indique que pratiquement toutes les variables étudiées interviennent dans la définition du collage, et les différents types de segmentation appliqués montrant que les sucres n'interviennent qu'en second plan.

Les résultats révèlent aussi qu'il est possible, par les méthodes de segmentation, de détecter les cotons très collants ou très peu collants d'une façon satisfaisante, mais, par contre, que pour les cotons moyennement collants, la détermination est beaucoup plus incertaine.

MOTS CLÉS : coton, fibre, collage, test minicarde, analyse des données, composantes principales, corrélations, maximisation variance intergroupe.

I. INTRODUCTION

L'incidence économique néfaste des cotons collants est actuellement très grande et de nombreux travaux de recherche sont effectués dans le monde entier. C'est ainsi que, dans certains pays africains, l'I.R.C.T., sous la conduite de J. MASSAT, procède à des expérimentations afin de mettre au point une méthode simple pour leur détection et connaître leur comportement en filature.

Aucune des méthodes couramment utilisées jusqu'à présent ne permet d'évaluer avec certitude le potentiel de collage d'un coton. On s'accorde à reconnaître la liaison entre le collage et la présence de miellats, substances sucrées sécrétées par certains insectes piqueurs parasites du cotonnier sans que cette liaison ait été établie d'une manière évidente.

Différentes méthodes chimiques sont utilisées : FEHLING, FOLIN, PERKINS, BENEDICT, étuvage, pH, extrait sec, chromatographie sur couche mince, en phases gazeuse ou

liquide à haute performance. Elles sont destinées à étudier, en particulier, les substances chimiques qui imprègnent ou souillent les fibres, mais ne suffisent pas, lorsqu'elles sont prises séparément, à expliquer le collage.

Un test mécanique, réalisé à l'aide d'une carte de micro-filature, donne, par contre, de bonnes indications sur le degré de collage : il est considéré comme assez sûr, bien que sa liaison avec le comportement de la fibre en filature industrielle n'ait pas été établie de manière rigoureuse. Ce test nécessite pour l'instant un investissement coûteux et demande une ambiance parfaitement contrôlée en température et humidité, condition qu'il est toujours difficile de maintenir stable en dehors d'un laboratoire très spécialisé.

L'objet du présent travail est de tenter d'expliquer le collage évalué à la minicarde par l'étude globale des tests et analyses chimiques et non par leur utilisation isolée.

II. MATÉRIEL ET MÉTHODES

1. Description de la population de cotons

58 cotons, provenant de cultivars et de pays différents, furent utilisés. Pour chacun de ces cotons, on a étudié :

— le degré de collage à la minicarde, selon la méthode mise au point par l'I.R.C.T. (1), le grade 1 correspondant à des cotons non collants, le grade 7 à des cotons très collants ;

— le pH qui caractérise l'acidité des fibres,

— le test de FEHLING qui indique la teneur en sucres réducteurs (2),

— la méthode BENEDICT qui évalue, sur les fibres elles-mêmes, les substances glucidiques réductrices (3),

— la colorimétrie, mesurée à l'aide du colorimètre SPIN-LAB mod. 154 ;

TABLEAU 1. — Caractéristiques de la population des 58 cotons analysés.

Sigle	Caractéristiques physiques ou chimiques	Valeurs		moyenne	écart-type	CV %
		minimum	maximum			
TCM	Test cardé	1	7	3,97	2,04	51,38
PH	pH	4,9	7,9	6,28	0,451	7,18
EST	Extrait sec total (%)	1,43	4,82	2,82	0,653	23,15
FEMA	Test FEHLING	0	9	3,39	2,77	81,71
TBE	Test BENEDICT	0	10	4,38	2,99	68,26
RDAET	Rd % avant étuvage	60,8	80,4	74,11	3,93	5,30
+ BAET	+ b avant étuvage	7,4	12,1	9,48	0,962	10,14
RDP	Rd % après étuvage	57,1	71,8	65,6	3,70	5,64
+ BP	+ b après étuvage	10,7	19,2	15,0	2,04	13,60
FRU	Fructose %	0	0,3	0,068	0,065	95,58
GLU	Glucose %	0,007	0,3	0,069	0,060	86,95
SAC	Saccharose %	0	0,18	0,023	0,034	147,82
MEL	Mélézitose %	0	0,03	0,0025	—	—
PH/EST	Rapport pH/extrait sec	1,22	5,10	2,39	0,775	32,42

a) valeurs Rd % (réflectance) et + b (degré de jaune) ;

b) les mêmes valeurs, après étuvage, suivant la méthode ELSNER (4).

— l'extrait sec total qui fournit le taux de substances organiques et minérales imprégnant les fibres (2).

— l'analyse chimique par chromatographie sur couche mince des sucres suivants (5) :

— fructose, glucose, saccharose, mélézitose.

Les résultats obtenus par cette dernière méthode ont été retenus de préférence à ceux donnés par les autres méthodes chromatographiques en phases gazeuse ou liquide, car ils présentent une meilleure corrélation avec le test de la minicarde.

Enfin, on a créé une variable pH/extrait sec total qui s'est révélée en assez forte corrélation avec le test minicarde.

Le tableau 1 résume les caractéristiques de la population analysée.

2. Description du matériel informatique

En raison du grand nombre d'observations et de variables à considérer, seuls les moyens informatiques pouvaient être utilisés et le travail qui va suivre a été réalisé à l'aide d'un micro-ordinateur ADDX et d'un mini-6 relié au gros calculateur du Centre National Universitaire Sud de Calcul, un ordinateur IBM 30-80.

Les logiciels utilisés sont, microstat pour l'ADDX, SPAD pour le mini-6, BMDP et NLT clotilde pour l'IBM 30-80.

3. Description des méthodes statistiques

L'« analyse des données » est un ensemble de méthodes aux fondements assez anciens (ADANSON, 1757 ; PEARSON, 1901 ; SPEARMAN, 1904) dont les buts sont la mise en évidence et la détermination des relations qui existent entre des individus, les paramètres qui les caractérisent et entre ces individus et leurs paramètres (6). Certains considèrent l'« A.D. » comme une nouvelle forme de statistiques descriptives (7) car toutes ces méthodes visent à la « description » de populations, cependant elles ne s'opposent pas aux techniques de statistiques classiques aux objectifs inférentiels (8) mais ont des concepts très différents. Ainsi, elles tentent de réduire au maximum la subjectivité et les a priori en restant le plus proche et le plus fidèle possible des données.

Le développement de l'utilisation de ces techniques suit celui de l'outil informatique et l'on peut aujourd'hui les subdiviser en deux groupes :

- les méthodes linéaires,
- les méthodes de classification automatique.

Les techniques utilisées pour ce travail appartiennent à ces deux ensembles.

A) L'analyse en composantes principales

C'est l'une des méthodes de base de l'analyse des données (9). Elle est très utilisée pour la description des tableaux de données quantitatives car elle permet de bien « visualiser » l'information qui y est contenue. Elle sert aussi pour classer un ensemble d'individus afin d'aboutir éventuellement à une partition de cet ensemble en classes d'éléments proches.

a) Principes

Pour cette méthode (fig. 1), on dispose d'un tableau « individus-caractères » ayant la forme suivante :

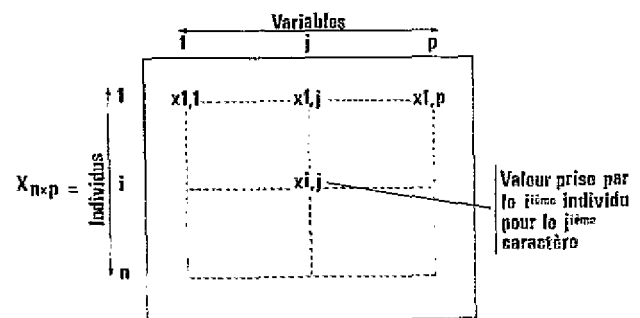


FIG 1.

S'il n'y a que deux variables V_1 et V_2 , il est facile de représenter dans un plan l'ensemble des données et de déterminer, en les caractérisant, les degrés de liaison entre individus ou groupe d'individus.

De la même façon, si l'on pouvait regarder dans un espace de dimension supérieure à 3, le classement des individus ne poserait pas plus de problèmes :

soit $B = \{ U_i / i = 1, \dots, p \}$.

Il suffit de placer les vecteurs de base U_i deux à deux perpendiculaires et à une même distance de l'origine 0 pour que les proximités entre individus soient mesurées par les

longueurs des segments les réunissant (8). Ainsi, dans cette représentation des n individus dans l'espace des p variables, les proximités des « points individus » s'interprètent immédiatement : leurs p coordonnées sont proches (10). De la même façon, on pourrait représenter les p variables dans l'espace des n individus et ici aussi les proximités des « points variables » proches indiqueraient que ces deux variables mesurent le même phénomène ou qu'elles sont liées par une relation particulière.

Malheureusement, nous ne sommes pas capables de « visualiser » des espaces de dimension supérieure à 3, et si l'on veut représenter le nuage des individus il faut se contenter de graphiques plans. Aussi ce que l'on voit n'est qu'une déformation de la réalité. Cette réduction entraîne des distorsions et le problème est, si les variables de départ sont liées, de passer à un tableau $Y_{n \times q}$ (individus-nouvelles variables), réduisant le nombre de variables tout en perdant le moins possible d'information.

b) Ajustements dans R_p et R_n

Pour passer de p à q variables ($q < p$) indépendantes, la technique consiste à projeter le nuage des individus sur un sous-espace affine W_q déterminé de telle sorte qu'on déforme le moins possible le nuage en projection. Si le sous-espace W_q est une droite, c'est la droite d'inertie minimum (fig. 2) qui sera la meilleure représentation du nuage des individus (6).

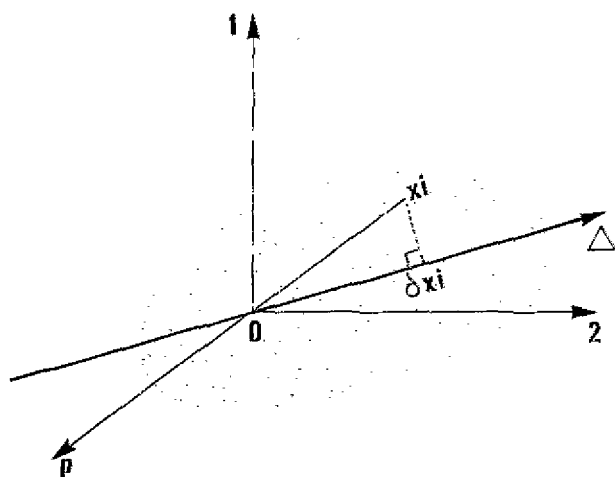


FIG. 2.

Le critère d'ajustement est celui des moindres carrés car on cherche à rendre minimum la somme des $d^2(X_i, \delta X_i)$. Or, d'après le théorème de Pythagore :

$$\sum d^2(0, X_i) = \sum d^2(0, \delta X_i) + \sum d^2(X_i, \delta X_i)$$

ou : $\sum d^2(X_i, \delta X_i) = \sum d^2(0, X_i) - \sum d^2(0, \delta X_i)$

comme : $\sum d^2(0, X_i)$ est déterminé, pour rendre minimum $\sum d^2(X_i, \delta X_i)$ il suffit de maximiser $\sum d^2(0, \delta X_i)$ (11). De la même manière, si l'on cherche le plan de projection qui conserve au mieux les distances du nuage en projection : les $d(X_i, X_j)$, il faudra rendre maximale, la somme des carrés de ces distances (9).

Comme les deux droites Δ^1 et Δ^2 sont perpendiculaires, on a : $\sum d^2(X_i, X_j) = \sum d^2(\delta^1 X_i, \delta^1 X_j) + \sum d^2(\delta^2 X_i, \delta^2 X_j)$ (th. de Pythagore). Il faut donc chercher Δ^1 puis Δ^2 en rendant maximum respectivement $\sum d^2(\delta^1 X_i, \delta^1 X_j)$ et $\sum d^2(\delta^2 X_i, \delta^2 X_j)$. De la même façon on peut déterminer Δ^3 , Δ^4 , ... Ce sont les principaux axes d'inertie du nuage, combinaisons linéaires des variables de départ.

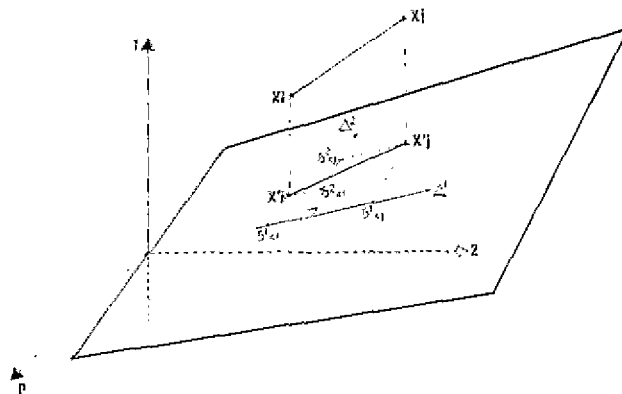


FIG. 3.

On a ainsi créé de nouvelles variables synthétiques (12) que l'on hiérarchise ; ce sont les composantes principales et la meilleure représentation du nuage des individus dans le sous-espace affine W_k se fera en prenant les K premières composantes principales.

B) La segmentation

C'est en fait une « famille » de techniques (13) dont le but est de réaliser à partir de variables qualitatives dites explicatives une partition d'une population qui explique au mieux une variable (ou des variables) qualitative ou quantitative dite « à expliquer » (14). Elles permettent de répondre à la question : comment décomposer de façon optimale la population étudiée en sous-populations discriminant bien un facteur donné ?

a) Principes

Il s'agit d'expliquer une variable Y , de nature quelconque (6) et la segmentation consiste à chercher d'abord la variable la plus explicative au sens d'un critère W , mesuré sur cette variable Y , de type suivant :

$$W(v) = \text{Max } \delta(P_1(v), P_2(v))$$

$$P(v) \in P_2(v)$$

où $P_2(v)$ est l'ensemble des partitions à deux classes que l'on peut obtenir sur la variable v et δ un indice d'agrégation (généralement construit à partir d'un indice de dissimilarité).

La variable la plus « explicative » de Y sera v^* tel que :

$$W(v^*) = \text{Max } W(v)$$

$$v \in V$$

où V est l'ensemble des variables explicatives.

b) Algorithme de segmentation « quantitative »

La procédure est la suivante : tout d'abord le programme va chercher, pour chacune des variables explicatives, l'ensemble des partitions en deux classes et il retiendra celle qui donne la meilleure valeur au critère, par exemple la variance interclasses, mesuré sur la variable à expliquer ; ou dit d'une autre façon, le programme détermine sur la population Ω , toutes les dichotomies définies par toutes les variables v , il calcule pour chacune d'elle un indicateur de liaison avec la variable à expliquer Y et il retiendra la variable et la dichotomie de ses modalités qui optimisent cet indicateur.

Une première étape permet de découper (ou segmenter) en deux sous-populations (Ω_1 et Ω_2) la population d'origine. Le programme recommence cette procédure sur les deux sous-populations et ainsi de suite. Il s'arrêtera en fonction de règles que nous lui aurons fixées : nombre de segments maxima, % de variance expliquée, etc. Le résultat peut être représenté par un arbre renversé ou couché.

« C'est l'analyste, en fonction des objectifs de l'étude et de la connaissance du phénomène étudié, qui choisit dans l'arbre la partition qui lui convient » (15).

A chaque nœud, on trouve une variable explicative et une dichotomie de ses modalités ainsi qu'un certain nombre d'informations : effectif du segment, % de variance expliquée, moyenne de la variable Y dans le segment, etc.

c) Segmentation sur coefficient de corrélation

Pour cette méthode, sensiblement différente de la précédente, l'algorithme est le suivant : la première étape consiste à déterminer la variable explicative la mieux corrélée avec la variable à expliquer. Dans la deuxième étape, on ordonne la population suivant la variable retenue et on la segmente en deux sous-populations d'effectifs équivalents ou presque. Les étapes suivantes sont des répétitions, sur

les sous-groupes formés, des deux premières étapes jusqu'à ce que les coefficients de corrélation ne soient plus significatifs. Ici aussi le résultat est présenté sous forme d'un arbre avec, à chaque nœud, un ensemble de critères caractérisant les sous-populations.

Malheureusement les résultats des segmentations restent assez pauvres (7) : à chaque itération ce n'est qu'un traitement mono-dimensionnel qui permet le découpage optimal en deux sous-populations, ou alors de deux variables explicatives très liées une seule sortira dans la segmentation l'autre restant masquée, etc. ; l'utilisateur doit donc constamment se référer aux valeurs prises par l'indicateur de liaison pour les différentes variables.

Pourtant la segmentation, malgré ses limites, est une alternative intéressante à la régression multiple et à l'analyse discriminante, dont elle reste par nature très différente.

III. ANALYSE EN COMPOSANTES PRINCIPALES (ACP)

L'analyse a porté sur 57 individus, un individu ayant été éliminé de la population initiale car présentant une valeur de pH trop éloignée de la moyenne. Sur les 20 variables explicatives de départ, seules ont été retenues celles qui étaient les mieux corrélées avec la variable Test Minicarde à expliquer et qui n'étaient pas très corrélées entre elles. La variable Fructose a également été éliminée car elle a un comportement identique à la variable Glucose et l'addition de ces 2 variables aurait trop « pesé » sur la constitution des axes. La variable Test Minicarde, ainsi que 6 individus moyens pour les différentes variables représentant chaque classe moyenne de collage à la minicarde (S1 à S6) ont été introduits en variables supplémentaires.

Les 8 variables retenues pour cette étude sont :

PH = pH
EST = Extrait sec total
TBE = Test BENEDICT
RDP = Réflectance % post étuvage
+ BP = Indice jaune post étuvage
GLU = Glucose
SAC = Saccharose
PH/EST = Rapport pH/Extrait sec total.

Cette dernière variable a été retenue car elle présente la meilleure corrélation avec le Test Minicarde (TCM).

Résultats et discussion

1. Diagonalisation de la matrice de corrélation

Les 3 premiers axes créés résument 84,4 % de l'information et à lui seul l'axe 1 en résume 59,8 %.

Le tableau 4 indique les vecteurs propres, coefficient des variables centrées réduites dans l'équation linéaire des axes principaux.

2. Etude des variables

Sur l'axe 1, toutes les variables sont bien représentées et particulièrement EST, TBE, + BP, SAC, PH/EST.

Sur l'axe 2, 4 variables sont assez bien représentées PH, GLU, SAC, PH/EST.

Sur l'axe 3, une seule variable est assez bien représentée, RDP.

Le tableau 5 indique les corrélations entre les différentes variables et les axes principaux.

TABEAU 2. — Matrice de corrélation sur 57 individus.

Variables	TCM	PH/EST	EST	PH	GLU	FRU	TBE	SAC	RDP	+ BP	FEMA
PH/EST	- 0,486										
EST	0,428	- 0,908									
PH	- 0,382	0,776	- 0,556								
GLU	0,360	- 0,881	0,365	- 0,199							
FRU	0,285	- 0,382	0,393	- 0,185	0,965						
TBE	0,271	- 0,594	0,600	- 0,446	0,629	0,654					
SAC	0,224	- 0,485	0,586	- 0,169	0,724	0,724	0,699				
RDP	- 0,236	0,525	- 0,583	0,242	- 0,320	- 0,317	- 0,473	- 0,438			
+ BP	0,215	- 0,656	0,642	- 0,441	0,654	0,701	0,699	0,631	- 0,537		
FEMA	0,240	- 0,669	0,684	- 0,478	0,630	0,671	0,712	0,678	- 0,587	0,794	
MEL	0,242	- 0,158	0,210	0,090	0,668	0,633	0,326	0,580	- 0,156	0,313	0,340

TABEAU 3.

	Axe 1	Axe 2	Axe 3	Axe 4
Valeurs propres (variance sur les axes principaux)	4,780 5	1,243 6	0,728 2	0,441 3
(%) Contribution à la variation totale	59,8	15,5	9,1	5,5
(%) Contribution cumulée	59,8	75,3	84,4	89,9

TABLEAU 4.

Variables	Vecteurs propres			
	Axe 1	Axe 2	Axe 3	Axe 4
PH	- 0,281 5	- 0,569 8	- 0,418 9	- 0,202 0
EST	0,394 3	0,249 0	- 0,204 5	- 0,426 5
TBE	0,383 9	- 0,187 0	0,179 7	- 0,033 1
RDP	- 0,303 8	0,004 6	0,782 2	- 0,386 0
+ BP	0,391 3	- 0,123 2	0,061 9	0,416 6
GLU	0,310 4	- 0,485 3	0,343 6	0,312 0
SAC	0,344 9	- 0,414 7	- 0,017 3	- 0,582 1
PH/EST	- 0,396 6	- 0,394 6	- 0,023 8	0,131 8

TABLEAU 5.

Variables	Composantes principales							
	Axe 1		Axe 2		Axe 3		Axe 4	
PH	- 0,6155	0,3788	- 0,6354	0,4038	- 0,3575	0,1278	- 0,1342	0,0180
EST	0,8622	0,7433	0,2776	0,0771	- 0,1745	0,0304	- 0,2833	0,0803
TBE	0,8394	0,7046	- 0,2086	0,0435	- 0,1534	0,0235	- 0,0220	0,0005
RDP	- 0,6642	0,4412	0,0052	0,0000	0,6761	0,4571	- 0,2564	0,0658
+ BP	0,8655	0,7318	- 0,1374	0,0189	0,0528	0,0028	0,2767	0,0766
GLU	0,6786	0,4605	- 0,5412	0,2929	0,2933	0,0860	0,2072	0,0429
SAC	0,7540	0,5686	- 0,4624	0,2139	- 0,0148	0,0002	- 0,3867	0,1495
PH/EST	- 0,8671	0,7518	- 0,4400	0,1936	- 0,0203	0,0004	0,0876	0,0077
Variable Supplémentaire TCM	0,4195	0,1760	0,1524	0,0232	0,0930	0,0087	- 0,0235	0,0006

1^{re} colonne : corrélations entre les variables et les axes principaux.2^e colonne : corrélations au carré.

3. Cercles de corrélations

— Plan 1-2 (fig. 4). Deux grands groupes sont situés sur l'axe 1 avec un pôle positif (EST, + BP, TBE, SAC, GLU) et un pôle négatif (RDP, PH/EST, TBE, PH). Sur l'axe 2, il y a un seul pôle négatif avec l'ensemble des variables sauf EST.

On note 3 grandes directions dans le plan 1-2 :

- une déterminée par SAC et GLU ;
- une déterminée par BP, TBE et qui s'opposent à RDP ;
- une déterminée par PH/EST et PH et qui s'opposent à EST.

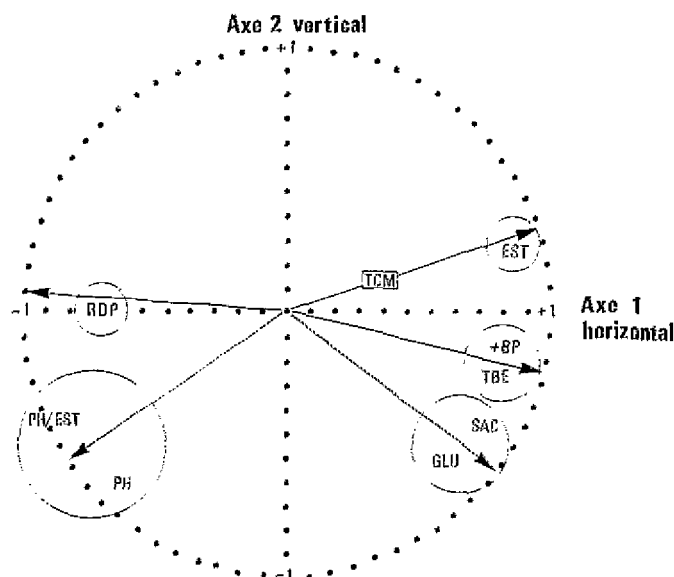


FIG. 4. — Cercle des corrélations plan 1-2.

FIG. 4. — Circle of correlations, surface 1-2.

- *Plan 1-3* (fig. 5) L'axe 3 est uniquement déterminé par RDP.
 Dans le plan 1-3, on note l'opposition entre PH/EST et PH avec l'ensemble des autres variables.

Ce plan n'apporte que peu de renseignements supplémentaires par rapport au plan 1-2.

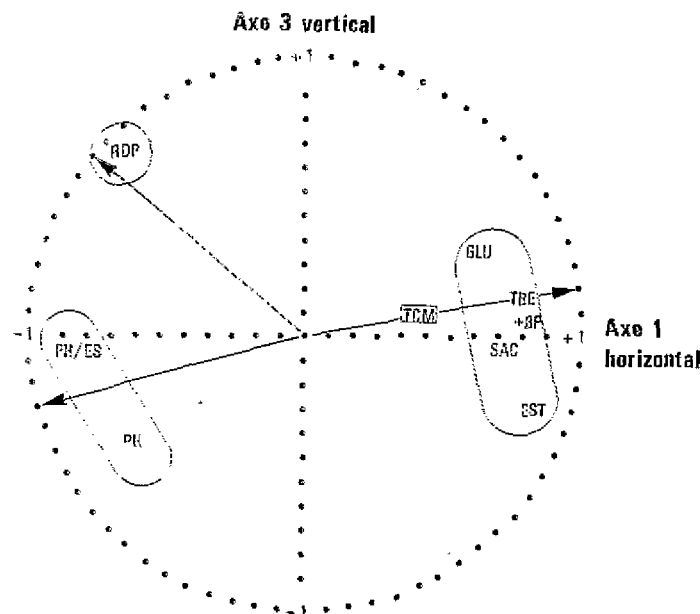


FIG. 5. — Cercle des corrélations plan 1-3.

FIG. 5. — Circle of correlations 1-3.

4. Etude des individus

Les individus sont représentés dans le plan 1-2 sur la figure 6. Il convient de noter qu'ils ont été classés par ordre croissant de collage et donc qu'aux numéros les plus faibles correspondent les cotons les moins collants et aux numéros les plus forts les cotons les plus collants. Sur cette même figure sont également représentés les individus supplémentaires qui sont aussi classés par ordre croissant de collage, S_1 étant le moins collant, S_6 le plus collant. Les coordonnées et la qualité de la représentation de ces derniers individus sont indiqués dans le tableau 6.

Sur la figure 6, on note un très net effet GUTTMAN : les individus se répartissent suivant un nuage en croissant parabolique où les extrêmes s'opposent. Ainsi le deuxième facteur est une fonction quadratique du premier. Nous nous trouvons donc devant un phénomène unidimen-

sionné. La ligne Y brisée obtenue en joignant les individus supplémentaires est donc proche de l'axe 1 et suit le croissant. Il semble bien que le phénomène caractérisé soit le collage.

Une analyse factorielle de correspondances multiples sur le tableau M_{ij} (avec i = l'ensemble des classes de collage et j l'ensemble des modalités des variables, les mieux représentées dans le plan 1-2 de l'A.C.P. et transformées en variables de classes permettrait de confirmer pleinement cette hypothèse.

Ceci permet de faire une partition en 3 groupes, cotons peu collants, moyennement collants, fortement collants.

L'axe 2 oppose les extrêmes aux moyens. On a donc du mal à caractériser ces derniers qui se trouvent au centre de la figure. Ceci se confirme par la proximité des individus supplémentaires S_2 , S_3 et S_4 .

TABEAU 6.

	Individus supplémentaires							
	Axe 1		Axe 2		Axe 3		Axe 4	
S1	- 2,0897	0,7610	- 1,0147	0,1794	- 0,3065	0,0164	0,3378	0,0199
S2	- 0,3035	0,2644	- 0,3275	0,3079	0,0850	0,0208	- 0,1931	0,1071
S3	- 0,2103	0,0991	0,4220	0,3991	0,3761	0,3169	0,1843	0,0761
S4	- 0,1673	0,0377	0,4617	0,2868	- 0,5063	0,3449	- 0,2352	0,0744
S5	0,6801	0,4979	0,6384	0,4387	- 0,0572	0,0035	- 0,1823	0,0358
S6	1,4611	0,8024	- 0,4507	0,0763	0,3962	0,0590	0,1733	0,0113

1^{ère} colonne : coordonnées des individus sur les axes principaux.

2^e colonne : cosinus carrés (qualité de la représentation).

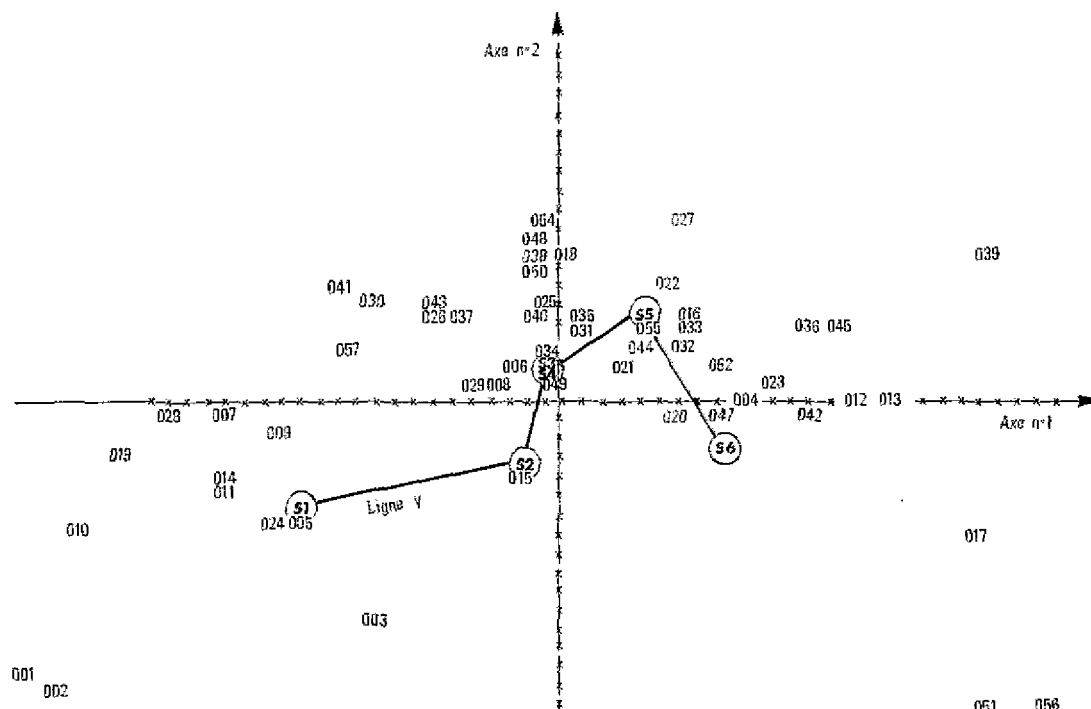


FIG. 6. — Représentation des individus dans le plan 1-2.

FIG. 6. — Representation of the individuals in plane 1-2.

Conclusions de l'A.C.P.

C'est sur l'axe 1 que la variable supplémentaire « test carde » est la mieux représentée.

Toutes les variables étudiées : sucres, tests indiquant la présence de sucres, extrait sec et pH sont bien représentées sur l'axe 1. On peut donc en déduire que l'axe 1 est celui du collage et que les variables qui contribuent à le créer, et

qui sont les mieux représentées, sont celles qui indiquent le mieux le potentiel de collage.

Cette étude montre qu'il sera assez facile de caractériser les grades extrêmes de collage (pas collants ou très collants) mais que, par contre, il y aura beaucoup d'incertitude pour l'identification des grades moyens. Elle semble également indiquer que l'appréciation du collage à la minicarde n'est peut-être pas très exacte et ceci justement pour les cotons moyennement collants.

IV. SEGMENTATIONS

1) Analyse par les corrélations

L'analyse a porté sur 58 individus. Pour segmenter, on retient la variable qui présente la meilleure corrélation avec le test de collage à la minicarde, on effectue un classement sur cette variable et l'on sépare la population en 2 groupes de valeurs de collage aussi différentes que possible. Pour chacun de ces groupes, on recherche ensuite la variable qui présente la meilleure corrélation avec le test collage à la minicarde et l'on continue la segmentation comme plus haut, jusqu'à ce que la corrélation avec le test carde ne soit plus significative.

Le tableau 7 donne la matrice de corrélation de la population de départ ainsi que les sigles retenus pour les différentes variables.

Le travail de segmentation a débuté à partir de 2 variables, le pH/extrait sec total et l'extrait sec total.

Résultats et discussion

Les figures 7 et 8 donnent, pour chacun des segments, les paramètres suivants :

- moyenne du collage : \overline{MC} (extrêmes des classes 1 à 7) ;
- nombre d'individus n :

— valeurs extrêmes de la variable qui détermine le segment ;

— coefficient de corrélation de la variable qui détermine le segment et sa signification.

TABLEAU 7.

Variables explicatives		Coefficient de corrélation avec la variable à expliquer (test minicarde)	Signification
pH/extrait sec	(PH/EST)	- 0,488	**
Extrait sec total	(EST)	0,432	**
pH	(PH)	- 0,376	**
Glucose	(GLU)	0,356	**
Fructose	(FRU)	0,281	*
Test de BENEDICT	(TBE)	0,269	*
Mélézitose couche mince	(MEL)	0,249	NS
Test de Fehling	(FEMA)	0,244	NS
Rd % post étuvage	(RDP)	- 0,234	NS
Saccharose	(SAC)	0,229	NS
+ b post étuvage	(+ BP)	0,206	NS
Rd % avant étuvage	(RDAET)	- 0,098	NS
+ b avant étuvage	(+ BAET)	0,094	NS

** Significatif 1 %, * significatif 5 %.

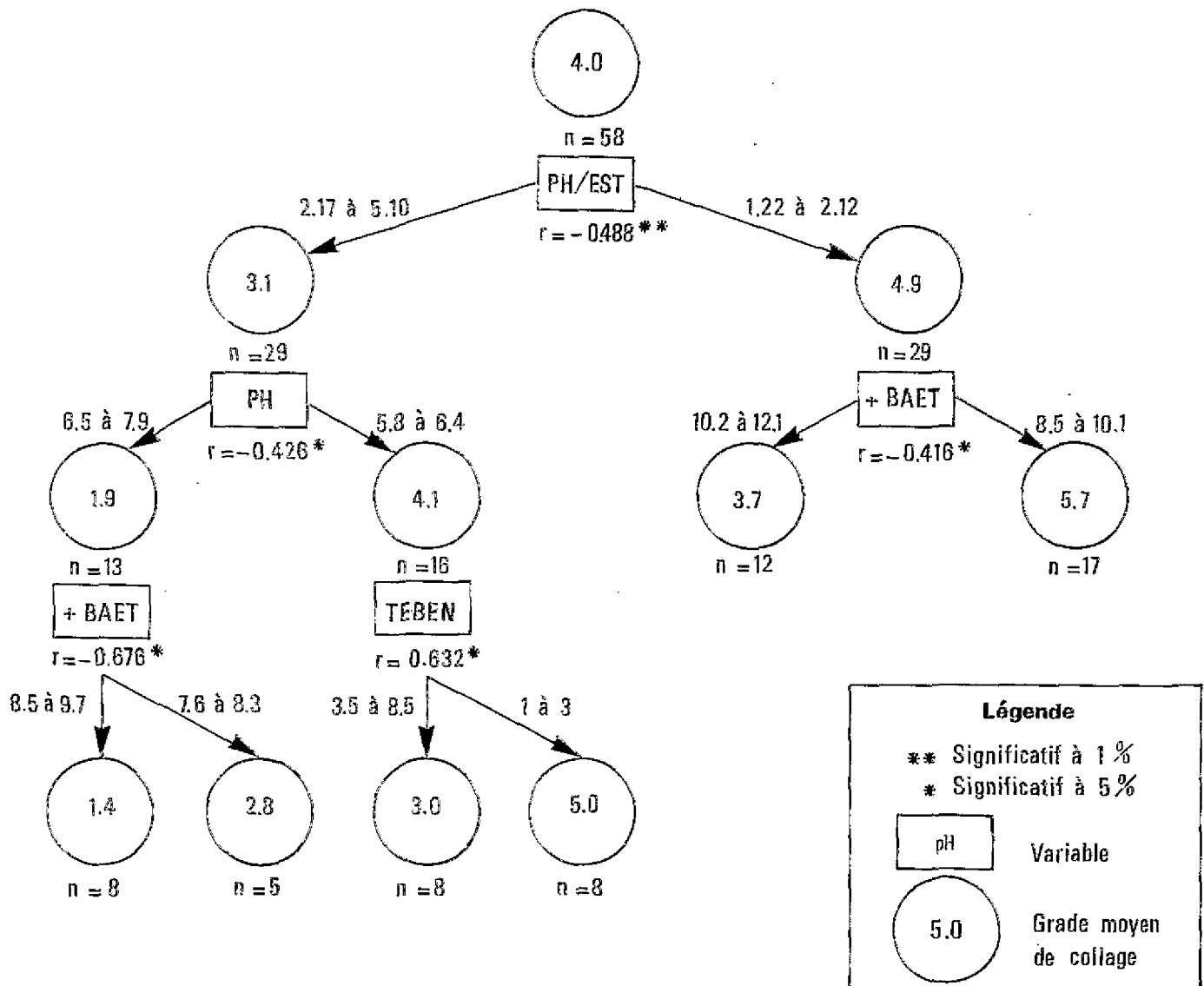


FIG. 7. — Analyse par les corrélations (segmentation) rapport pH/extrait sec total.

FIG. 7. — Analysis by correlations (segmentation) pH/total dry extract ratio.

Si la valeur du rapport est la plus forte (2,17 à 5,10) on a affaire aux cotons les moins collants ($\overline{MC} = 3,1$) et dans ce cas c'est le pH qui permet une nouvelle segmentation. Les cotons à pH les plus faibles (5,8 à 6,4) sont les plus collants ($\overline{MC} = 4,1$) et les cotons à pH les plus forts (6,5 à 7,9) sont les moins collants ($\overline{MC} = 1,9$). Dans ce dernier groupe, c'est le degré de jaune + b (+ BAET) qui permet de segmenter ; les cotons de + b le plus faible (7,6 à 8,3) sont les plus collants ($\overline{MC} = 2,8$) ; ceux de + b le plus élevé (8,5 à 9,7) sont les moins collants ($\overline{MC} = 1,4$). Dans le groupe de pH faible (5,8 à 6,4), c'est le test de BENEDICT qui permet de segmenter, pour les valeurs élevées (3,5 à 8,5) les cotons sont les moins collants ($\overline{MC} = 3,0$) et pour les valeurs les plus faibles (1 à 3) on a les cotons les plus collants ($\overline{MC} = 5,0$). Il convient de noter ici cette apparente contradiction puisque le test de BENEDICT indique les substances glucidiques réductrices.

Si la valeur du rapport est la plus faible (1,22 à 2,12), on a affaire aux cotons les plus collants ($\overline{MC} = 4,9$) et, dans ce cas, c'est le degré de jaune + b (+ BAET) qui permet de segmenter. Les cotons à + b faible (8,5 à 10,1) sont les plus collants ($\overline{MC} = 5,7$) et les cotons à + b élevé (10,2 à 12,1) sont les moins collants ($\overline{MC} = 3,7$).

Lorsque le point de départ de la segmentation est la deuxième variable la mieux corrélée avec le collage à la minicarde, c'est-à-dire l'extrait sec total, deux groupes peuvent être définis. Si l'extrait sec est faible (1,43 à 2,81), on a des cotons faiblement collants ($\overline{MC} = 3,0$). S'il est plus élevé (2,88 à 4,82), on a des cotons nettement plus collants : ($\overline{MC} = 4,9$). Dans le groupe des cotons à extrait sec élevé (2,88 à 4,82), c'est l'indice de jaune + b (+ BAET) qui permet une nouvelle segmentation. Si cet indice est faible (8,5 à 10,1), on a affaire aux cotons les plus collants ($\overline{MC} = 5,6$) ; s'il est plus fort (10,2 à 12,1), les cotons sont moins collants ($\overline{MC} = 3,7$).

Dans le groupe des cotons à faible extrait sec 1,43 à 2,82 c'est la variable pH qui permet une nouvelle segmentation. Lorsque les valeurs de pH sont les plus faibles — 5,8 à 6,4 — il s'agit des cotons les plus collants : 3,9 ; si les valeurs de pH sont plus élevées — 6,5 à 7,9 — les cotons sont moins collants : 1,9. Pour les cotons à faible pH, c'est le test de BENEDICT qui permet une nouvelle segmentation. Pour un test de BENEDICT de 1 à 5 on a des cotons collants : 4,6 ; de 6 à 8,5 on a des cotons peu collants : 2,0. Pour les cotons de pH élevé, c'est le degré de jaune + b qui permet de segmenter. Un + b de 7,6 à 8,3 donne les

cotons les plus collants : 2,8 et une + b de 8,5 à 9,7 donne les cotons les moins collants : 1,4.

TABLEAU 8.

Segmentation à partir de pH/extrait sec (fig. 7)			Segmentation à partir d'extrait sec total (fig. 8)		
Valeur moyenne du collage dans les segments	Ecart type	nb d'individus	Valeur moyenne du collage dans les segments	Ecart type	nb d'individus
5,7	1,57	17	5,6	1,74	18
5,0	1,41	8	4,6	1,69	11
3,7	1,49	12	3,7	1,49	12
3,0	1,92	8	2,8	0,84	5
2,8	0,83	5	2,0	1,15	4
1,4	0,51	8	1,4	0,51	8

Conclusions de l'analyse par les corrélations

Il convient de remarquer qu'aussi bien dans un schéma que dans l'autre, ce sont les variables extrait sec total et pH ou leur combinaison qui sont les mieux corrélées avec le test à la minicarde et donc qui détectent le mieux les cotons collants. Ensuite, on trouve le test de BENEDICT qui indique la présence de sucres réducteurs et le + b qu'il est difficile d'expliquer ; les cotons les moins colorés et donc les plus ternes seraient les plus collants. Les sucres eux-mêmes n'apparaissent pas dans ces schémas.

Peut-on appliquer cette méthode à la détection des cotons collants ? Certainement pour les cotons extrêmes (très collants ou pas collants), mais il restera une grande part d'incertitude pour les cotons classés moyennement collants.

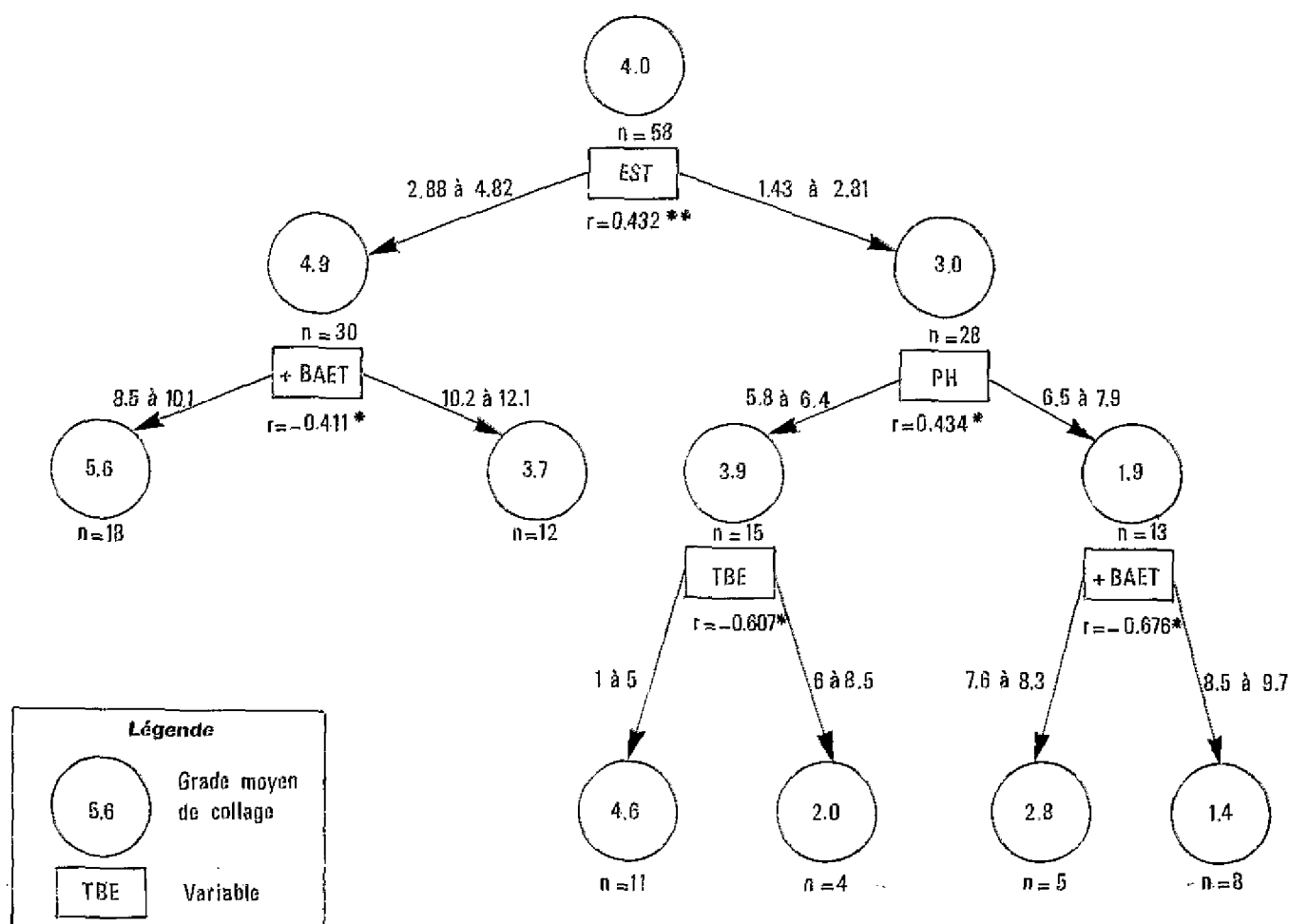


FIG. 8. — Analyse par les corrélations (segmentation) extrait sec total.

FIG. 8. — Analysis by correlations (segmentation) total dry extract.

V. SEGMENTATION QUANTITATIVE PAR MAXIMISATION DE L'ÉCART DE VARIANCE INTER-GROUPE

L'analyse a porté sur 57 individus et l'on a retenu les variables suivantes :

Réflectance après étuvage Rd %	(RDP)
Degré de jaune + b après étuvage	(+ BP)
Extrait sec total	(EST)
pH	(PH)
Fructose	(FRU)
Glucose	(GLU)
Saccharose	(SAC)
Mélezitose	(MEL)
Test BENEDICT	(TBE)
Test FEHLING	(FEMA)

Pour cette étude, les classes de collage vont de 0 à 6. Le tableau 9 donne le pourcentage de la variance « collage » expliquée par chacune des variables.

TABLEAU 9.

Variables explicatives	Pourcentage de la variance expliquée
pH	14,69 %
Extrait sec total	14,43 %
Glucose	12,94 %
Saccharose	8,32 %
Test BENEDICT	7,74 %
+ b après étuvage	4,21 %
Rd % après étuvage	4,15 %
Test FEHLING	2,79 %
Fructose	2,33 %
Mélezitose	1,67 %

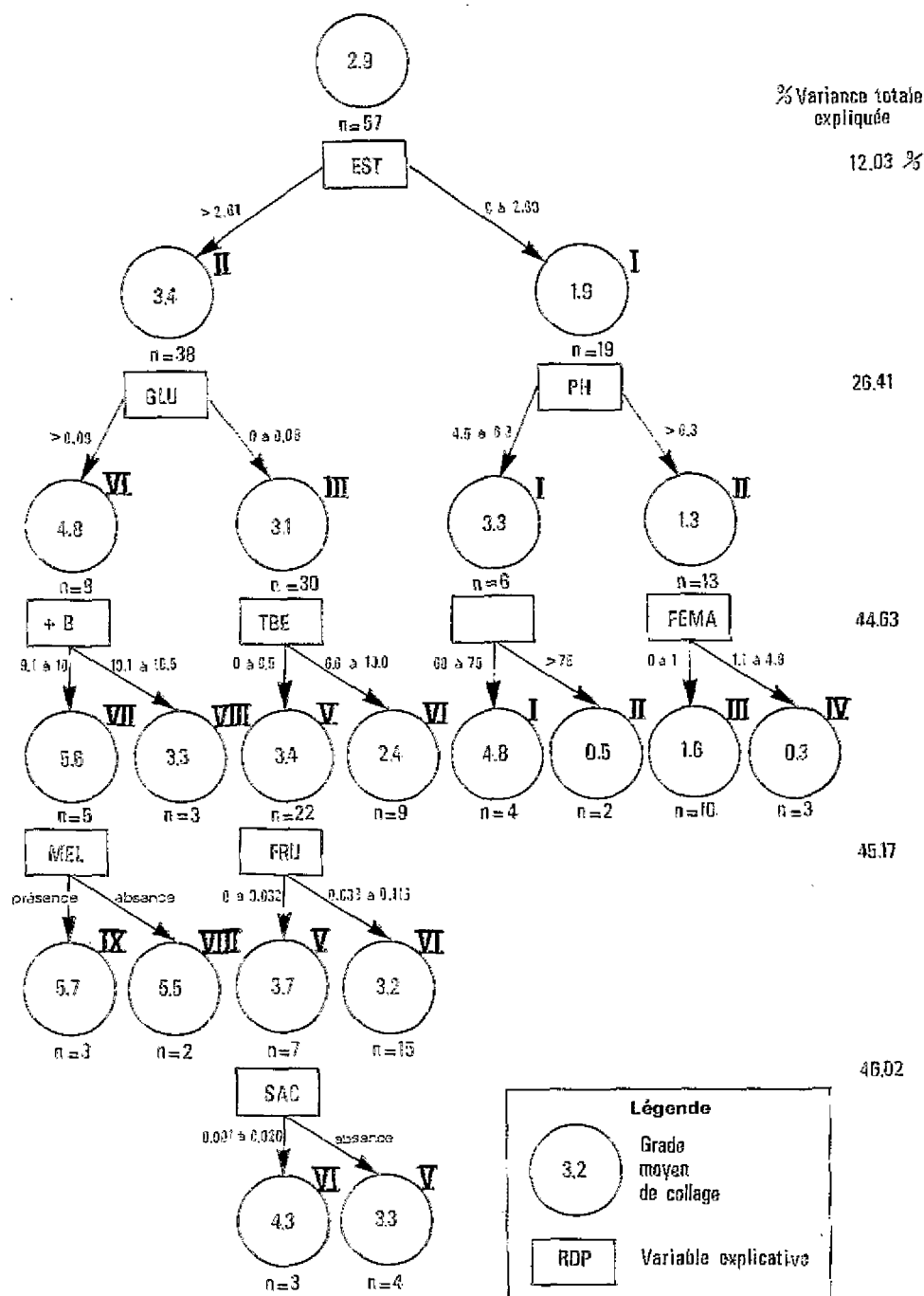


FIG. 9. — Segmentation quantitative.

FIG. 9. — Quantitative segmentation.

Résultats et discussion

Dans la figure 9, on indique pour chacun des segments :

- la moyenne de collage (de 0 à 6),
- le nombre d'individus n ,
- les valeurs extrêmes de la variable qui détermine le segment,
- le pourcentage de la variance « collage » expliquée à chaque niveau de segmentation.

Le tableau 10 indique le nombre de segments obtenus à chaque niveau d'itération, la variance expliquée et la signification du test F.

TABLEAU 10.

Numéro d'itération	Nombre de segments	Variance expliquée		Test F
		valeur	%	
1	2	0,500	12,03	7,52**
2	4	1,097	26,41	6,34**
3	8	1,850	44,63	5,64**
4	4	1,880	45,17	4,30**
5	2	1,910	46,02	3,92**
		4,155	100,00	

La variable qui permet de segmenter au départ est l'extrait sec total : les cotons les plus collants ($\overline{MC} = 3,4$) ont un extrait sec supérieur à 2,61 ; les cotons les moins collants ($\overline{MC} = 1,9$) ont un extrait sec variant de 0 à 2,60.

Pour les cotons les plus collants, c'est ensuite le taux de glucose qui permet de continuer la segmentation : pour un taux $< 0,08$ % le collage moyen est de 3,1 et pour un taux $> 0,08$ % il est de 4,8.

Pour les cotons les moins collants, c'est le pH qui permet de segmenter. De 4,5 à 6,3 de pH, le collage moyen est de 3,3 ; au-dessus de 6,3 de pH, le collage moyen est de 1,3.

Ensuite, sans donner de détails, la segmentation peut se poursuivre par les tests indiquant la présence de sucres. BENEDICT et FEHLING et colorimétrie Rd % et + b.

En dernier lieu, ce sont les sucres mélézitose, fructose et saccharose, qui permettent de terminer la segmentation.

Il convient enfin de remarquer qu'il y a des contradictions apparentes à partir du niveau 2, les valeurs les plus faibles du test BENEDICT donnent les cotons les plus collants ; il en est de même pour le test de FEHLING.

Le tableau 11 indique les valeurs moyennes de collage et

TABLEAU 11.

Niveaux d'itération	Valeur moyenne du collage dans les segments	Ecart type	Nombre d'individus
0	2,9	2,04	57
1	3,4	1,8	38
	1,9	2,1	19
2	4,8	1,6	8
	3,3	1,7	6
	3,1	1,7	30
	1,3	1,6	13
	5,6	0,5	5
3	4,8	1,1	4
	3,4	1,6	22
	3,3	1,9	3
	2,4	1,9	9
	1,6	1,7	10
	0,5	0,5	2
	0,3	0,5	3
	5,7	0,5	3
	5,5	0,5	2
	3,7	1,7	7
4	3,2	1,5	15
	4,3	1,7	3
	3,3	1,6	4

les écarts-types des segments aux différents niveaux d'itération.

On constate dans ce tableau que si l'on arrive à obtenir des groupes de collage très différents les uns des autres, les écarts-types sont élevés.

Conclusions de l'analyse par maximisation de l'écart des variances inter-groupe

Comme pour la segmentation avec l'aide des corrélations, les variables qui permettent d'effectuer les premières segmentations sont le pH et l'extrait sec total avec aussi, dans ce cas, le glucose. Viennent ensuite les tests de colorimétrie après étuvage et les tests indiquant la présence de sucres, qui se comportent d'ailleurs de façon inverse à ce que l'on aurait pu supposer. Les sucres eux-mêmes n'interviennent qu'en dernier lieu et les groupes de collage qu'ils permettent de distinguer diffèrent très peu entre eux.

Comme dans le cas précédent, on arrivera assez bien à distinguer par cette méthode les cotons très collants de ceux qui ne collent pas, mais il y aura une très grande part d'incertitude pour identifier les cotons moyennement collants.

VI. DISCUSSION ET CONCLUSIONS GÉNÉRALES

La matrice de corrélation, tableau 12, indique que les variables qui sont le plus fortement liées avec le collage, mesuré à l'aide d'une minicarde, sont l'extrait sec total (EST) et le pH (PH) puis les sucres, fructose (FRU) et glucose (GLU) et, enfin, les tests de FEHLING (FEMA) et de BENEDICT (TBE).

La matrice de corrélation indique également que le pH et l'extrait sec sont bien liés entre eux, que le pH n'est pas du tout lié avec les sucres, que l'extrait sec total est lié avec les sucres, et qu'en colorimétrie c'est surtout l'indice de jaune après étuvage (+ BP) qui est lié avec les sucres.

En conséquence, le collage n'est pas seulement dû aux sucres contrairement à ce que l'on pensait.

Ces constatations sont confirmées par l'observation des résultats de l'analyse en composantes principales. Toutes les variables contribuant à créer l'axe 1 (axe du collage) et qui sont les mieux représentées : pH, extrait sec total,

sucres, tests FEHLING et BENEDICT, colorimétrie après étuvage, interviennent dans la définition du collage à la minicarde.

Les segmentations à partir des corrélations et de la maximisation de l'écart de variance intergroupe montrent que les variables les plus importantes sont l'extrait sec et le pH. Les sucres arrivent en tout dernier.

Est-il possible de détecter les cotons collants par les méthodes de segmentation étudiées dans ce travail ? Certainement pour les extrêmes (cotons très collants ou pas du tout collants), mais il reste une incertitude quant au degré de collage réel des cotons classés moyennement collants.

Il semble, d'autre part, que les seuils des valeurs des variables définissant chaque segment mériteraient d'être mieux précisés, ce qui ne peut se faire que par l'étude d'un grand nombre d'échantillons et très probablement par la

participation de plusieurs laboratoires. Enfin l'appréciation du collage à la minicarde ne paraît pas être valable pour les degrés moyens ; ceci apparaît clairement dans l'analyse en

composantes principales avec les individus supplémentaires dont chacun est la moyenne des variables d'un degré de collage.

TABLEAU 12.

	TCM	PH	EST	FEMA	TBE	RDAET	+ BAET	RDP	+ BP	FRU	GLU	SAC
PH	- 0,376											
EST	0,432	- 0,572										
FEMA	0,244	- 0,467	0,685									
TBE	0,269	- 0,401	0,587	0,708								
RDAET	- 0,098	0,002	- 0,120	0,060	0,097							
+ BAET	0,094	- 0,399	0,669	0,615	0,475	- 0,346						
RDP	- 0,234	0,212	- 0,570	- 0,584	- 0,473	0,565	- 0,684					
+ BP	0,208	- 0,355	0,609	0,778	0,696	0,316	0,591	- 0,536				
FRU	0,281	- 0,164	0,379	0,665	0,654	0,288	0,273	- 0,318	0,700			
GLU	0,356	- 0,165	0,352	0,625	0,629	0,246	0,265	- 0,321	0,654	0,966		
SAC	0,229	- 0,197	0,592	0,680	0,693	- 0,004	0,417	- 0,433	0,580	0,715	0,651	
MEL	0,249	- 0,073	0,258	0,350	0,297	0,043	0,082	- 0,138	0,249	0,574	0,606	0,576

Sign : 5 % = 0,250 ; 1 % = 0,325.

RÉFÉRENCES BIBLIOGRAPHIQUES

1. MASSAT (J.), 1981. — Note Interne. Mise au point d'un test de collage à la minicarde. *Div. de technol. I.R.C.T.*
2. BOURELY (J.), 1980. — Contribution à l'étude des sucres du cotonnier. *Cot. Fib. trop.*, 35, 2, 189-208.
3. RIMON (D.). — Chemical methods for the evaluation of stickiness in Cotton Fibers *Agr. Res. Org. Inst. Volcani Center, Bet Dagan 50-250, Israel.*
4. ELSNER (O.), 1982. — Detecting Sugars in Cotton Lint *Testing Research Journal*, 538.
5. BOURELY (J.), GUTKNECHT (J.), FOURNIER (J.), 1984. — Etude chimique du collage des fibres du coton. 1^{re} partie : Rôle des sucres et de miellats dans le processus du collage. *Cot. Fib. trop.*, 39, 3, 47-53.
6. DIDAY (E.), LEMAIRE (L.), FOUGET (J.), TESTU (F.), 1982. — Eléments d'analyse de données *Dunod.*
7. FENELON (J.-P.), 1981. — Qu'est-ce que l'analyse de données ? *Lefonen.*
8. CAILLEZ (F.), PAGES (J.-P.), 1976. — Introduction à l'analyse des données. *Editions SMASH.*
9. BOUROCHE (J.-M.), SAPORTA (G.), 1983a. — L'analyse des données. *P.U.F., Que sais-je ?*
10. LE BART (L.), MORINEAU (A.), FENELON (J.-P.), 1977. — Technique de la description statistique *DUNOD.*
11. MAURE (F.), RENOU (A.), VAISSAYRE (M.), 1980. — Approche multidimensionnelle d'un ensemble d'observations phytosanitaires par l'analyse en composantes principales. *Cot. Fib. trop.*, 35, 2.
12. JOLY (A.), 1980. — Note de recherche.
13. HUGUES (M.), 1970. — Segmentation et typologie. *Bordas Management.*
14. BACCINI, 1984. — Exposé au Séminaire de Statistique du 8/3/1984 au C.N.U.S.C. Montpellier.
15. Analyses des données et A.P.L., Rapport C.E.A. R 4753, 1976.

Chemical analysis of stickiness in cotton

Part 2 : Application of data analysis to a population of 58 cottons

J. Fournier *, J. Gutknecht **, E. Jallas *** et J. Bourelly ****

* Technologist.

** Head of the Technology Division.

*** Biometrician.

**** Head of the Laboratory of Textile Plant Chemistry, I.R.C.T.-C.I.R.A.D., B.P. 5035, Montpellier, Cedex, France.

SUMMARY

As to explain cotton stickiness which undoubtedly hinders the proper operation of spinning mills, 58 cottons with different degrees of stickiness measured by the minicard test have been examined by data analysis : principal components, segmentation

by correlations and segmentation by maximization of the variance difference between groups.

The explanatory variables selected for the analysis are the following : phi, total dry extract after fiber maceration in water, pre

and post-heating colour measurement and sugar content by thin-layer chromatography.

The results obtained show that stickiness and presence of sugars should not be systematically associated, since the variables that are the most correlated with stickiness are pH and total dry extract. pH is not correlated at all with the presence of sugars and dry extract is only slightly correlated.

Principal component analysis indicates that almost all the variables examined are involved in the definition of stickiness; the different types of segmentation applied show that the role of sugars is secondary.

Also, the results show that segmentation methods can satisfactorily detect very sticky or unsticky cottons. They are much less reliable for moderately sticky cottons.

KEY WORDS : cotton, fibre, stickiness, minicard test, data analysis, principal components, correlations, maximization of the variance between groups.

I. INTRODUCTION

Sticky cottons have a very negative economic incidence at the moment and many research works are under way throughout the world. For instance, in several African countries, I.R.C.T., under the supervision of Mr. MASSAT, is performing experiments to develop a simple method to detect them and know their behaviour during spinning.

None of the methods currently used so far can estimate with certainty the potential stickiness of one cotton. It is admitted that stickiness is connected with the presence of honeydew, which is a sugar substance secreted by several piercing insect pests of cotton, but this has never been clearly established.

Various chemical methods are used : FERLING, FOLIN, PERKINS, BENEDICT, heating, pH, dry extract, thin-layer high performance gas or liquid chromatography. They are

meant to study the chemical substances that impregnate or dirty the fibers. But, they cannot, taken separately, explain stickiness.

On the opposite, a mechanical test carried out with a Shirley miniature card is a good indicator of the degree of stickiness; it is considered fairly reliable although its correlation with fiber behaviour during industrial spinning has not been rigorously established. For the time being, this test requires high investments. Also, a perfectly controlled environment regarding temperature and moisture is essential, being this condition very difficult to maintain stable outside a highly specialized laboratory.

This paper aims at explaining stickiness, estimated with a minicard test, by the overall study, and not isolated use of chemical tests and analyses.

II. MATERIALS AND METHODS

1. Description of the cotton population

58 cottons from various cultivars and countries have been examined. For each cotton, we have measured the following characteristics :

- the degree of stickiness after a minicard test, according to the method developed by I.R.C.T. (1). Grade 1 corresponds to unsticky cottons and grade 7 to very sticky cottons,
- pH, which characterizes fiber acidity,
- FERLING test, which indicates the reducing sugar content (2),
- BENEDICT method, which estimates, on the fibers themselves, the reducing glucidic substances (3),
- colour measurement, with a SPINLAB 154 colorimeter,
- reflectance (grayness) value (Rd%) and yellowness degree (+ b),

— reflectance value and yellowness degree, after heating, according to ELSNER method (4),

— total dry extract, which gives the amount of mineral and organic substances impregnating the fibers (2),

— chemical analysis, by thin layer chromatography, of the following sugars : fructose, glucose, saccharose, melezitose.

The results obtained with this method have been selected in preference to those given by the other gas or liquid chromatography methods, since they are better correlated with the minicard test.

Lastly, a variable pH/total dry extract has been created, which proved fairly strongly correlated with the minicard test.

The characteristics of the population analysed are summed up in table 1.

TABLE 1. — Characteristics of the cotton population analyzed.

Initials	Physical or chemical characteristics	Values		mean	Standard deviation	CV %
		minimum	maximum			
MCT	Minicard test	1	7	3.97	2.04	51.38
PH	pH	4.9	7.9	6.28	0.451	7.13
TDE	Total Dry Extract %	1.43	4.82	2.82	0.853	23.15
FEMA	FERLING test	0	9	3.39	2.77	81.71
BET	BENEDICT test	0	10	4.33	2.99	68.26
PHRD	Pre-heating Rd %	60.3	80.4	74.11	3.93	5.30
+ PHB	Pre-heating + b	7.4	12.1	9.48	0.962	10.14
RDP	Post-heating Rd %	57.1	71.8	65.6	3.70	5.64
+ BP	Post-heating + b	10.7	19.2	15.0	2.04	13.60
FRU	Fructose %	0	0.3	0.068	0.065	95.58
GLU	Glucose %	0.007	0.3	0.069	0.060	86.95
SAC	Saccharose %	0	0.18	0.023	0.034	147.82
MEL	Melezitose %	0	0.03	0.0025	—	—
PH/EST	pH/dry extract ratio	1.22	5.10	2.39	0.775	32.42

2. Description of the computers

Owing to the large number of observations and variables to consider, only computers could be used. The present study has been carried out with an ADDX micro-computer and a mini-6 connected with the IBM 30-80 of the C.N.U.S.C. (Centre National Universitaire Sud de Calcul).

The following software was used : microstat for ADDX, SPAD for the mini-6, BMDP and NLT clouide for the IBM 30-80.

3. Description of the statistical methods

The « data analysis » groups several methods which were developed quite a long time ago (ADANSON, 1757 ; PEARSON, 1901 ; SPEARMAN, 1904). It aims at emphasizing and determining the relations which exist between individuals, the parameters characterizing them and between these individuals and their parameters (6). « D.A. » is sometimes regarded as a new form of descriptive statistics (7) since all these methods aim at « describing » populations. However, they are not opposed to conventional statistical techniques with inferential objectives (8) but have very different concepts. They attempt a maximum reduction in subjectivity and apriorism by remaining as close and faithful as possible to the data.

The increasing use of these techniques follows that of computerized means and today, they can be divided into two groups :

- linear methods,
- automatic classification methods.

The techniques used in this work belong to these groups.

A) Principal component analysis

It is one of the basic methods of data analysis (9). It is commonly used to describe tables of quantitative data since the information contained is well displayed. It is also used to classify a group of individuals to lead perhaps to a partition of this group into classes of close elements.

a) Principles

We have a table « individuals characters » with the shape given in fig. 1.

If there are only two variables V1 and V2, it is easy to represent in a plane all the data and to determine, characterizing them, the degrees of correlation between individuals or groups of individuals.

Similarly, if we could look at a space of a dimension over 3, it would be as easy to classify the individuals :

$$\text{Let } B = \{ U_i / i = 1, \dots, p \}.$$

It will be enough if we place the basic vectors U_i perpendicular each to each at an equal distance from the origin O to measure the proximities between individuals by the lengths of the segments joining them (8). On this representation of n individuals in the space of p variables, proximities of « points individuals » are immediately interpreted : their p co-ordinates are close (10). Similarly, the p variables could be represented in the space of the n individuals. Also, proximities of close « points variables » would indicate that these two variables measure the same phenomenon or are connected by a particular correlation.

Unfortunately, we cannot « visualize » spaces of a dimension over 3, and if we want to represent the cloud of individuals, we have to content ourselves with graphical surfaces. What we see is only a deformation of reality. This reduction leads to distortions and if the original variables are correlated, the problem is to lead to a table

$Y_i \times q$ (individuals-new variables), reducing the number of variables and losing as little information as possible.

b) Adjustments in R_p and R_n

In order to pass from p to q independent variables ($q < p$), the technique consists in projecting the cloud of individuals on an affine sub-space W_q determined in such a way that the cloud projected is as little deformed as possible. If the sub-space W_q is a line, the minimum inertia line (fig. 2) will be the best representation of the cloud of individuals (6).

The criterion for adjustment is that of least squares since we want to reduce to a minimum the sum of $d^2(X_i, \delta X_i)$. According to Pythagoras' theorem :

$$\Sigma d^2(0, X_i) = \Sigma d^2(0, \delta X_i) + \Sigma d^2(X_i, \delta X_i)$$

$$\text{or } \Sigma d^2(X_i, \delta X_i) = \Sigma d^2(0, X_i) - \Sigma d^2(0, \delta X_i)$$

as $\Sigma d^2(0, X_i)$ is determined to make $\Sigma d^2(X_i, \delta X_i)$ minimum, maximizing $\Sigma d^2(0, \delta X_i)$ will be enough (11).

Similarly, if we are looking for the projection plan that best keeps the distances of the cloud projected, the sum of these squared distances will be made maximal.

As the two lines Δ_1 and Δ_2 are perpendicular, we have $\Sigma d^2(X_i, X_j) = \Sigma d^2(\delta^1 x_i, \delta^1 x_j) + \Sigma d^2(\delta^2 x_i, \delta^2 x_j)$ (Pythagoras' theorem). Therefore, we have to seek Δ_1 then Δ_2 by making maximum $\Sigma d^2(\delta^1 x_i, \delta^1 x_j)$ and $\Sigma d^2(\delta^2 x_i, \delta^2 x_j)$ respectively. Δ_3 and Δ_4 (etc.) can be determined this way. They are the main inertia axes of the cloud, linear combinations of the original variables. New synthetic variables are thus created (12) and organized into a hierarchy ; they are the principal components and the best representation of the cloud of individuals in the affine sub-space W_k will be obtained by taking the k first components.

B) Segmentation

It is in fact a technique « family » which aims at achieving, from so-called explanatory qualitative variables, the partition of a population which best explains one (or several) quantitative or qualitative variables (s) called « to be explained » (14). They make it possible to answer the following question : how can the population examined be divided in an optimal way into sub-populations differentiating properly a given factor.

a) Principles

The point is to explain a variable Y , of any nature (6), and segmentation consists in seeking first the most explanatory variable on the basis of a criterion W , measured on this variable of the following type :

$$W(v) = \text{Max } \delta(P_1(v), P_2(v)) \\ P(v) \in P_2(v)$$

where $P_2(v)$ is the group of partitions with two classes that can be obtained on the variable v and δ an aggregation index (generally constructed from a dissimilarity index).

The most « explanatory » variable of Y will be v^* so that

$$W(v^*) = \text{Max } W(v) \\ v \in V$$

where V represents all the explanatory variables.

b) Algorithm of « quantitative » segmentation

The method is the following : for each explanatory variable, the programme will first seek all the partitions into two classes and select the one which gives the best value to the criterion, for instance the variance between classes, measured on the variable to be explained. In other words, the programme determines on the population Ω all the dichotomies defined by all the variables v . It calculates for each of them an indicator of correlation with the variable to be explained Y and will select the variable and dichotomy of its modalities which optimize this indicator.

A first stage allows the original population to be divided (or segmented) into two sub-populations, Ω_1 and Ω_2 . The programme repeats this operation on the two sub-populations and so on. It will stop according to the rules laid down by us : maximum number of segments, % of variance explained, etc. The result can be represented by an inverted or lying tree. « According to the objectives of the study and the knowledge of the phenomenon examined, the analyst will select in the tree the partition that suits him » (15).

Each knot contains an explanatory variable and a dichotomy of its modalities together with several pieces of information : number of individuals in the segment, % of variance explained, mean of the variable Y in the segment, etc.

c) Segmentation on correlation coefficient

For this method, which is quite different from the previous one, the algorithm is as follows : the first stage consists in determining the explanatory variable which is

the best correlated with the variable to be explained. In the second stage, the population is organized according to the variable selected and segmented into two sub-populations with the same, or almost the same, number of individuals. The next stages are repetitions, on the sub-groups formed, of the first stages until correlation coefficients become insignificant. Here again, the result is presented as a tree where each knot contains a group of criteria characterizing the sub-populations.

Unfortunately, the results of the segmentations are relatively poor (7) : each iteration is only a univariate analysis which leads to the optimal division into two sub-populations ; or, out of two highly correlated explanatory variables, only one will be visible in the segmentation, the other being hidden, etc. The user must therefore constantly refer to the values taken by the indicator of correlation for the different variables.

However, despite its limits, segmentation is an interesting alternative to multiple regression and discriminant analysis from which it is very different by nature.

III. PRINCIPAL COMPONENT ANALYSIS (PCA)

57 individuals were analysed : one was eliminated from the original population since its pH value was too far from the average. Out of the 20 original explanatory variables, we have only selected those which were the best correlated with the variable Minicard test to be explained and were not correlated between themselves. The variable Fructose was also eliminated since its behaviour is identical to that of Glucose and the addition of these two variables would have had an excessive influence on the formation of the axes. The variable Minicard test as well as 6 individuals, means of the different variables, representing each class of mean stickiness values after the Minicard test (S1 to S6) have been introduced as additional variables.

The 8 variables selected for this study are the following :

PH : pH,
TDE : total dry extract,
BET : BENEDICT test,
RDP : post-heating reflectance %,
+ BP : post-heating yellowness degree,
GLU : glucose,
SAC : saccharose,
PH/TDE : pH/total dry extract ratio.

The latter has been selected since it has the best correlation with the Minicard test (MCT).

Results and discussion

1. Diagonalization of the matrix of correlation

The first 3 axes created sum up 84.4 % of the information, and 59.8 % are summed up in axis 1 alone.

2. Study of the variables

All the variables are well represented on axis 1, and especially TDE, BET, +BP, SAC and PH/TDE.

Four variables are relatively well represented on axis 2 : PH, GLU, SAC and PH/TDE.

Only one variable is relatively well represented on axis 3 : RDP.

Table 5 shows the correlations between the different variables and the principal axes.

TABLE 2. — Matrix of correlation on 57 individuals.

Variables	MCT	PH/TDE	TDE	PH	GLU	FRU	BET	SAC	RDP	+ BP	FEMA
PH/TDE	- 0.485										
TDE	0.428	- 0.908									
PH	- 0.382	0.776	- 0.556								
GLU	0.360	- 0.881	0.366	- 0.199							
FRU	0.285	- 0.382	0.393	- 0.185	0.965						
BET	0.271	- 0.594	0.600	- 0.446	0.629	0.654					
SAC	0.224	- 0.465	0.586	- 0.169	0.724	0.724	0.699				
RDP	- 0.236	0.525	- 0.583	0.242	- 0.320	- 0.317	- 0.473	- 0.438			
+ BP	0.215	- 0.656	0.642	- 0.441	0.654	0.701	0.699	0.601	- 0.537		
FEMA	0.240	- 0.669	0.684	- 0.478	0.630	0.671	0.712	0.678	- 0.587	0.794	
MEL	0.242	- 0.158	0.210	0.090	0.668	0.633	0.326	0.580	- 0.156	0.313	0.340

TABLE 3.

	Axis 1	Axis 2	Axis 3	Axis 4
Latent values (variance on principal axes)	4.7805	1.2436	0.7282	0.441 3
(%) Contribution to total variation	59.8	15.5	9.1	5.5
(%) Pooled contribution	59.8	75.3	84.4	89.9

TABLE 4.

Variables	Latent vectors			
	Axis 1	Axis 2	Axis 3	Axis 4
PH	- 0.2815	- 0.5698	- 0.4189	- 0.2020
TDE	0.3943	0.2490	- 0.2045	- 0.4265
BET	0.3839	- 0.1870	0.1797	- 0.0331
RDP	- 0.3038	0.0046	0.7922	- 0.3860
+ BP	0.3913	- 0.1232	0.0619	0.4166
GLU	0.3104	- 0.4853	0.3436	0.3120
SAC	0.3449	- 0.4147	- 0.0173	- 0.5821
PH/TDE	- 0.3966	- 0.3946	- 0.0238	0.1318

3. Circles of correlations

Surface 1-2 (fig. 4). Two major groups are located on axis 1 with a positive pole (TDE, +BP, BET, SAC, GLU) and a negative pole (RDP, PH/TDE, PH).

On axis 2, there is only one negative pole with all the variables except TDE.

There are three main directions in surface 1-2 :

- one determined by SAC and GLU,
- one determined by +BP, BET and opposed to RDP,
- one determined by PH/TDE and opposed to TDE.

Surface 1-3 (fig. 5). Axis 3 is determined by RDP only. In surface 1-3, PH/TDE and PH are opposed to all the other variables.

As compared to surface 1-2, this surface only brings few additional information.

4. Study of the individuals

The individuals are represented in surface 1-2 on fig. 6. It should be noted that they have been classified in

increasing order of stickiness. In other words, the lowest numbers correspond to the least sticky cottons and the highest numbers to the most sticky cottons. This figure also shows the additional individuals which are classified in increasing order of stickiness as well, being S1 the least sticky and S6 the most sticky. The co-ordinates and quality of representation of these individuals are shown in table 6.

A marked GUTTMAN effect is visible on figure 6 : the individuals are distributed according to a parabolic crescent-shaped cloud where extremes are opposed. The second factor is thus a quadratic function of the first. It is therefore a univariate phenomenon.

The broken line y obtained by joining the additional individuals is close to axis 1 and follows the crescent. Apparently, the phenomenon characterized is stickiness.

A factor analysis of multiple correspondences on the matrix $M_{i,j}$ (with i : all the classes of stickiness and j : all the modalities of the variables which are the best represented on the surface 1-2 of PCA and transformed into class variables) would allow this hypothesis to be fully confirmed.

TABLE 5.

Variables	Principal components							
	Axis 1		Axis 2		Axis 3		Axis 4	
PH	- 0.6155	0.3788	- 0.6354	0.4038	- 0.3575	0.1278	- 0.1342	0.0180
TDE	0.8622	0.7433	0.2776	0.0771	- 0.1745	0.0304	- 0.2833	0.0803
BET	0.8394	0.7046	- 0.2086	0.0435	- 0.1534	0.0235	- 0.0220	0.0005
RDP	- 0.8642	0.4412	0.0052	0.0000	0.6761	0.4571	- 0.2564	0.0658
+ BP	0.8555	0.7318	- 0.1374	0.0189	0.0528	0.0028	0.2767	0.0766
GLU	0.6786	0.4603	- 0.5412	0.2929	0.2933	0.0860	0.2072	0.0429
SAC	0.7540	0.6686	- 0.4624	0.2139	- 0.0148	0.0002	- 0.3867	0.1495
PH/TDE	- 0.8671	0.7518	- 0.4400	0.1936	- 0.0203	0.0004	0.0876	0.0077
Additional variables								
MCT	0.4195	0.1760	0.1524	0.0232	0.0930	0.0087	- 0.0235	0.0006

1st column : correlations between variables and principal axes.
2nd column : squared correlations.

TABLE 6.

	Additional individuals							
	Axis 1		Axis 2		Axis 3		Axis 4	
S1	- 2.0897	0.7610	- 1.0147	0.1794	- 0.3065	0.0164	0.3378	0.0199
S2	- 0.3035	0.2644	- 0.3275	0.3079	0.0850	0.0208	- 0.1931	0.1071
S3	- 0.2103	0.0991	0.4220	0.3991	0.3761	0.3169	0.1843	0.0761
S4	- 0.1673	0.0377	0.4617	0.2868	- 0.5063	0.3449	- 0.2352	0.0744
S5	0.6801	0.4979	0.6384	0.4387	- 0.0572	0.0035	- 0.1823	0.0358
S6	1.4611	0.8024	- 0.4507	0.0763	0.3962	0.0590	0.1733	0.0113

1st column : co-ordinates of the individuals on the principal axes.
2nd column : squared cosinus (quality of representation).

A partition into 3 groups is possible : unsticky or slightly sticky cottons, moderately sticky cottons and very sticky cottons.

Axis 2 opposes extremes to moderates. It is therefore difficult to characterize the latter which are in the centre of the figure. This is confirmed by the proximity of the additional individuals S2, S3 and S4.

Conclusions of the MCA

It is on axis 1 that the additional variable « Minicard

test » is best represented. All the variables examined (sugars, tests indicating the presence of sugars, dry extract and pH) are properly represented on axis 1. It can be deduced that axis 1 is the axis of stickiness and that the variables contributing to its creation, which are the best represented, are those which best indicate the potential stickiness. This study shows it will be fairly easy to characterize the extreme grades of stickiness (unsticky or very sticky). On the other hand, the identification of moderate grades will be very uncertain. Also, it shows that the estimate of stickiness by a minicard test may not be very exact, precisely for moderately sticky cottons.

IV. SEGMENTATIONS

1) Analysis by correlations

The analysis was made on 58 individuals. In order to segment, we have selected the variable which presented the best correlation with the minicard test ; a classification was made on this variable and the population was divided into two groups of stickiness values as different as possible. For each group, we have found the variable which presented the best correlation with the minicard test and we continued to segment as above, until the correlation with the minicard test became insignificant.

Table 7 gives the matrix of correlation of the original population as well as the initials selected for each variable.

TABLE 7.

Explanatory variables		Coefficient of correlation with the variable to explain (minicard test)	Significance
pH/dry extract	(PH/TDE)	- 0.488	**
Total dry extract	(TDE)	0.432	**
pH	(PH)	- 0.376	**
Glucose	(GLU)	0.356	**
Fructose	(FRU)	0.281	*
BENEDICT test	(BET)	0.269	*
Melazitose	(MEL)	0.249	NS
FENLING test	(FEMA)	0.244	NS
Post-heating Rd %	(RDP)	- 0.234	NS
Saccharose	(SAC)	0.229	NS
Post-heating + b	(+ BP)	0.206	NS
Pre-heating Rd %	(PHRD)	- 0.098	NS
Pre-heating + b	(+ PHB)	0.094	NS

** Significant 1 %, * significant 5 %.

Segmentation started from 2 variables : pH/total dry extract and total dry extract.

Results and discussion

For each segment, figure 7 and 8 give the following parameters :

- mean stickiness (extremes of classes 1 through 77),
- number of individuals n,
- extreme values of the variable that determines the segment,
- coefficient of correlation of the variable which determines the segment and its significance.

If the value of the ratio is the highest (2.17 to 5.10), we are dealing with the least sticky cottons (mean stickiness :

3.1) ; and in that case, a new segmentation is possible thanks to pH. The cottons with the lowest pH (5.8 to 6.4) are the most sticky (mean stickiness : 4.1) and the cottons with the highest pH (6.5 to 7.9) are the least sticky (mean stickiness : 1.9). In that group, segmentation is possible thanks to yellowness degree +b (+PHB) ; the cottons with the lowest +b (7.6 to 8.3) are the most sticky (mean stickiness : 2.8) ; those with the highest +b (8.5 to 9.7) are the least sticky (mean stickiness : 1.4). In the group with the lowest pH (5.8 to 6.4) segmentation is possible thanks to BENEDICT test. For the highest results (3.5 to 8.5) the cottons are the least sticky (mean stickiness : 3.0) and for the lowest values (1 to 3) the cottons are the most sticky (mean stickiness : 5.0). This apparent contradiction should be noted here since the BENEDICT test indicates the reducing glucidic substances.

If the value of the ratio is the lowest (1.22 to 2.12), we are dealing with the most sticky cottons (mean stickiness : 4.9) ; in that case, segmentation is possible thanks to the yellowness degree +b (+PHB). The cottons with a low +b (8.5 to 10.1) are the most sticky (mean stickiness : 5.7) and those with a high +b (10.2 to 12.1) are the least sticky (mean stickiness : 3.7).

When the starting point of segmentation is the second variable which is the best correlated with the Minicard test, that is total dry extract, two groups can be defined. If dry extract is low (1.43 to 2.81), the cottons are slightly sticky (mean stickiness : 3.0). If it is higher (2.88 to 4.82), the cottons are markedly more sticky (mean stickiness : 4.9). In the group of cottons with a high dry extract (2.88 to 4.82), segmentation is possible thanks to the yellowness degree +b (+PHB). If the degree is low (8.5 to 10.1), we are dealing with the most sticky cottons (mean stickiness : 5.6) ; if it is higher (10.2 to 12.1), the cottons are less sticky (mean stickiness : 3.7).

In the group of cottons with a low dry extract (1.43 to 2.82), segmentation is possible thanks to the variable pH. When the values of pH are the lowest (5.8 to 6.4), the cottons are the most sticky (mean stickiness : 3.9) ; if the values of pH are the highest (6.5 to 7.9), the cottons are less sticky (mean stickiness : 1.9). For the cottons with a low pH, segmentation is possible thanks to the BENEDICT test. When the BENEDICT test ranges from 1 to 5, the cottons are sticky (mean stickiness : 4.6) ; from 6 to 8.5, the cottons are slightly sticky (mean stickiness : 2.0). For the cottons with a high pH, segmentation is possible thanks to the yellowness degree +b (+PBH). When +b is 7.6 to 8.3, the cottons are the most sticky (mean stickiness 2.8), and when it is 8.5 to 9.7, the cottons are the least sticky (mean stickiness : 1.4).

This method allowed us to form groups of cottons with different degrees of stickiness. But we should admit however that the standard deviation is rather high, as shown in table 8.

TABLE 8.

Segmentation from pH/dry extract (fig. 7)			Segmentation from total dry extract (fig. 8)		
Mean value of stickiness in the segments	Standard deviation	Number of individuals	Mean value of stickiness in the segments	Standard deviation	Number of individuals
5.7	1.57	17	5.6	1.74	18
5.0	1.41	8	4.8	1.69	11
3.7	1.49	12	3.7	1.49	12
3.0	1.92	8	2.8	0.84	5
2.8	0.83	5	2.0	1.15	4
1.4	0.51	8	1.4	0.51	3

V. QUANTITATIVE SEGMENTATION BY MAXIMIZATION OF THE VARIANCE DIFFERENCE BETWEEN GROUPS

The analysis was made on 57 individuals and we have selected the following variables :

- post-heating reflectance Rd% (RDP),
- post-heating yellowness degree +b (+BP),
- total dry extract (TDE),
- pH (PH),
- Fructose (FRU),
- Glucose (GLU),
- Saccharose (SAC),
- Melezitose (MEL),
- BENEDICT test (BENT),
- FEHLING test (FEMA).

For this study, the classes of stickiness range from 0 to 6. Table 9 gives the percentage of the variance « stickiness » explained by each variable.

TABLE 9.

Explanatory variables	Percentage of the variance explained
pH	14.69 %
Total dry extract	14.43 %
Glucose	12.34 %
Saccharose	3.32 %
BENEDICT test	7.74 %
Post-heating + b	4.21 %
Post-heating Rd %	4.15 %
FEHLING test	2.79 %
Fructose	2.33 %
Melezitose	1.57 %

Results and discussion

In figure 9 we indicate for each segment :

- the mean stickiness (from 0 to 6),
- the number of individuals n,
- the extreme values of the variable determining the segment,
- the percentage of the variance « stickiness » explained at each level of segmentation.

Table 10 shows the number of segments obtained at each level of iteration, the variance explained and the significance of test F.

The variance that makes segmentation possible at the beginning is total dry extract : the most sticky cottons (mean stickiness : 3.4) have a dry extract of over 2.61 ; the

Conclusions of the analysis by correlations

It should be noted that in both cases, the variables total dry extract and pH or their combinations are the best correlated with the Minicard test. They are therefore the most capable of detecting sticky cottons. We find next the BENEDICT test which indicates the presence of reducing sugars and +b which is difficult to explain ; the least coloured cottons, therefore the duldest, would be the most sticky. The sugars themselves do not appear.

Can this method be used to detect sticky cottons ? It is certainly possible for extreme cottons (very sticky or unsticky) but largely uncertain for moderately sticky cottons.

TABLE 10.

Iteration number	Number of segments	Variance explained		Test F
		Value	%	
1	2	0.500	12.03	7.52**
2	4	1.097	28.41	6.34**
3	8	1.850	44.63	5.64**
4	4	1.880	45.17	4.30**
5	2	1.910	46.02	3.92**
		4.155	100.00	

least sticky cottons (mean stickiness : 1.9) have a dry extract ranging from 0 to 2.60.

For the most sticky cottons, segmentation continues thanks to the glucose content ; when this content is below 0.08 %, mean stickiness is 3.1 ; when it is over 0.08 % mean stickiness is 4.8.

For the least sticky cottons, segmentation is possible thanks to pH. When pH ranges from 4.5 to 6.3, mean stickiness is 3.3 ; when pH is over 6.3, mean stickiness is 1.3.

Without entering into details, we shall say that segmentation can be continued by the tests indicating the presence of sugars (BENEDICT, FEHLING and colour measurements Rd% and +b). Ultimately, segmentation is finished thanks to the sugars (melezitose, fructose and saccharose). It should be noted that there are apparent contradictions as from level 2 : the lowest values of the BENEDICT test give the most sticky cottons ; the same applies to the FEHLING test.

Table 11 indicates the mean values of stickiness and the standard deviations of the segments at the different levels of iteration.

This table shows that while the groups of stickiness we obtain are very different the ones from the others, standard deviations are high.

Conclusions of the analysis by maximization of the variance difference between groups

As in the case of segmentation by correlations, the variables that make the first segmentations possible are pH and total dry extract as well as, in this case, glucose. Next, we find post-heating colour measurement and the tests indicating the presence of sugars, the behaviour of which is

TABLE 11.

Levels of iteration	Mean value of stickiness in the segments	Standard deviation	Number of individuals
0	2.9	2.04	57
1	3.4	1.8	38
	1.9	2.1	19
2	4.8	1.6	8
	3.3	1.7	6
	3.1	1.7	30
	1.3	1.6	13
3	5.6	0.5	5
	4.8	1.1	4
	3.4	1.6	22
	3.3	1.9	3
	2.4	1.9	9
	1.6	1.7	10
	0.5	0.5	2
	0.3	0.5	3
4	5.7	0.5	3
	5.5	0.5	2
	3.7	1.7	7
	3.2	1.5	15
5	4.3	1.7	3
	3.3	1.6	4

opposite what could be expected. The sugars themselves are involved lastly and the groups of stickiness they allow to distinguish do not differ very much from the others.

As in the previous case, it will be fairly easy to distinguish by this method very sticky cottons from unsticky cottons. But the identification of moderately sticky cottons will be largely uncertain.

VI. GENERAL DISCUSSION AND CONCLUSIONS

The matrix of correlation (table 12) shows that the variables which are the most correlated with stickiness, measured by a Minicard test, are first total dry extract and pH (PH) then sugars (fructose (FRU)) and glucose (GLU) and lastly FEHLING (FEMA) and BENEDICT tests (BET).

The matrix of correlation also shows that pH and dry extract are well correlated between themselves. Besides, it indicates that pH is not correlated at all with the sugars and that total dry extract is correlated with the sugars. It also shows that, regarding colorimetry, it is mainly the yellowness degree that is correlated with the sugars.

Consequently, stickiness is not only due to sugars, contrary to what was supposed before.

These remarks are confirmed by the observation of the results of the principal component analysis. All the variables which contribute to create axis 1 (axis of stickiness) and are the best represented (pH, total dry extract, sugars, FEHLING and BENEDICT tests, post-heating colour measurement), are involved in the definition of stickiness by a Minicard test.

Segmentations by correlations and maximization of the variance difference between groups show that the most important variables are dry extract and pH. Sugars are the least important ones.

TABLE 12.

	MCT	PH	TDE	FEMA	BET	PHRD	+ PHB	RDP	- BP	FRU	GLU	SAC
PH	- 0.376											
TDE	0.432	- 0.572										
FEMA	0.244	- 0.467	0.685									
BET	0.269	- 0.401	0.587	0.703								
PHRD	- 0.098	0.002	- 0.120	0.060	0.087							
+ PHB	0.094	- 0.399	0.569	0.615	0.475	- 0.346						
RDP	- 0.234	0.212	- 0.570	- 0.584	- 0.473	0.565	- 0.684					
- BP	0.206	- 0.355	0.608	0.778	0.896	0.316	0.591	- 0.536				
FRU	0.281	- 0.154	0.379	0.665	0.654	0.288	0.273	- 0.318	0.700			
GLU	0.356	- 0.165	0.352	0.825	0.629	0.246	0.265	- 0.321	0.654	0.986		
SAC	0.229	- 0.197	0.592	0.680	0.893	- 0.004	0.417	- 0.433	0.580	0.715	0.651	
MEL	0.249	- 0.073	0.258	0.350	0.297	0.043	0.082	- 0.138	0.249	0.574	0.606	0.576

Sign : 5 % = 0.250, 1 % = 0.325.

Can sticky cottons be detected by the segmentation methods reviewed in this paper ? It is certainly possible for extreme cottons (very sticky or unsticky), but the determination of the real stickiness degree of moderately sticky cottons will be uncertain.

Besides, it seems that the values of the variables defining each segment should be pinpointed. This can be achieved

only by studying many samples and probably with the participation of several laboratories. Lastly, the estimate of stickiness by a Minicard test does not seem applicable to moderate degrees ; this is clearly shown in the principal component analysis with the additional individuals which are each the mean value of the variables of a degree of stickiness.

RESUMEN

Para explicar la pegadura de las fibras de algodón, cuyo efecto perjudicial a la buena marcha de las hilanderías es seguro, 58 algodones de diferentes grados de pegadura medidos por la minicarda fueron estudiados por el análisis de datos : componentes principales, segmentación por correlaciones y segmentación por maximización de la diferencia de varianza entre grupos.

Las variables explicativas seleccionadas para el análisis son las siguientes : pH, extracto seco total tras la maceración de la fibra en agua, colorimetría antes y después calentamiento, y contenido en varios azúcares por cromatografía sobre capa fina.

Los resultados obtenidos indican primero que se tiene que evitar de confundir sistemáticamente la pegadura con la presencia

de azúcares, ya que las variables más correlacionadas con la pegadura son el pH y el extracto seco total. El pH no está nada correlacionado con la presencia de azúcares y el extracto seco total lo es ligeramente.

El análisis en componentes principales indica que casi todas las variables estudiadas intervienen en la definición de la pegadura y los diferentes tipos de segmentación aplicados muestran que el papel de los azúcares es secundario.

También, los resultados indican que se puede, con los métodos de segmentación, descubrir los algodones muy pegajosos o muy poco pegajosos de una forma satisfactoria. En cambio, la determinación de los algodones medianamente pegajosos es mucho menos segura.